# Supplementary Materials

## Two waves of *de novo* methylation during mouse germ cell development

Antoine Molaro, Ilaria Falciatori, Emily Hodges, Alexei A. Aravin, Krista Marran, Shahin Rafii, W. Richard McCombie, Andrew D. Smith, and Gregory J. Hannon

correspondence to: hannon@cshl.edu

**This file includes:**

**Supplementary Figure Legends**

**Supplementary Figure 1.** Browser snapshot of a piRNA dependent, constitutitvely hypomethylated and piRNA independent L1 copy. (A) piRNA targeted L1Md_A found on chr8 (plus strand). The bottom tracks show the single CpG methylation levels in Mili (-/-) animals (green) and WT animals in (red). Methylation is low over the promoter (grey box) in mutant animals and high in WT. (B) Example of a constitutively hypomethylated L1Md_F3 found on chr7 (minus strand). The promoter displays low average methylation in both genotypes (grey box). (C) Example of a piRNA-independent L1Md_F found on chr14 (plus strand). Promoter methylation is high in both genotypes (grey box).

**Supplementary Figure 2**. (A) Total number of promoter CpGs of untrucated L1 elements (as in Fig. 2B) as a function of differential methylation in Mili (-/-) spermatocytes (Mut-WT). (B) Average size of piRNA-dependent (black), constitutively hypomethylated (red) or piRNA-independent (grey) retrotransposon copies in a subset of LINE and LTR sub-families. Internal sequences (-int) and LTR sequences (_LTR) are plotted separately for LTR sub-families.

**Supplementary Figure 3.** Distribution of promoter average methylation for all three categories of L1 (A) and ETn (B) retrotransposon copies at E16.5 (Seisenberger et al. 2012). Insertions on chr9 were chosen for analysis and display.

**Supplementary Figure 4.** qRT-PCR quantification of L1 and IAP RNA levels. (A) Relative abundance of retrotransposon transcripts in E13.5 PGCs compared to somatic cells. Fold changes and standard error across technical triplicates (black line) were computed from ΔΔCts using either MapK1 (left) or SDHA (right) as normalizer genes (van den Bergen et al. 2009). Primers sets over L1_ORF2 and the 5'UTRs of L1Md_T and _A were used to probe L1s. Primers against the internal sequence of IAPEz were used for IAPs. (B) Relative abundance of the same transcripts as in (A) in E16.5 PGCs compared to E13.5 PGCs.

**Supplementary Figure 5**. Model of retrotransposon *de novo* methylation during germ cell reprogramming. 4 groups of retrotransposons can be described: piRNA-dependent, piRNA-independent, constitutively hypomethylated and hypermethylated. In E13.5 PGCs (left panel) genome-wide DNA methylation is at its lowest. Only a few retrotransposon copies (mostly from the IAP family) retain high levels of DNA methylation at their regulatory region (red dots, Hypermethyalted, 4$^{th}$ group). Past E13.5 (middle panel), a wave of non-selective, default, *de novo* methylation is initiated; piRNA-independent copies (~90% of all genomic insertions) begin to regain most of their promoter methylation (red dots, 2$^{nd}$ group). piRNA-dependent and constitutively hypomethylated copies (less than ~10% of all genomic insertions, 1$^{st}$ and 3$^{rd}$ group) remain hypomethylated. Finally between E16.5 and the spermatocyte stage (right panel), the piRNA pathway is adaptively programmed to target young and active copies (piRNA-dependent, red dots, 1$^{st}$ group). Constitutively hypomethylated copies never regain DNA methylation at their promoters.

## Supplementary Materials and Methods

### Read Mapping

*bisulfite sequencing*: Reads were mapped as described previously (Molaro et al. 2011) to the mm9 reference genome assembly. *RNA-seq (long and small RNAs)*: The RMAP (Smith et al. 2009) aligner was used to map reads to the mm9 reference genome, allowing up to 5 mismatches or 6% of the read length. Reads were mapped to the whole mm9 genome to capture reads mapping to annotated repeats. For multiple mappers one location was selected uniformly at random.

### Calling HMRs and DMRs in spermatocyte libraries

Hypomethylated regions (HMRs) were identified as described previously (Molaro et al. 2011) and differentially methylated regions were determined as non-overlapping portions of HMRs between two methylomes, filtered to require at least 5 significantly differing CpG sites per interval using the method described in (Hodges et al. 2011).

### Repeat Annotations

Repeat annotations were taken from the RepeatMasker track downloaded through the UCSC Table Browser with the following modification. Copies of repeats were merged if they appeared consecutively on the same strand, and if the portion of the repeat consensus that matched the annotation in the genome were immediately consecutive in the consensus. This modification was to rebuild fragmented annotations for which the genome contains a single actual repeat insertion. However, solo LTRs were kept as individual elements.

Annotations

*HMR and DMRs*: HMR and DMR coordinate were intersected with UCSC annotation tracks. Annotation tracks were prioritized as follows: repeats, genes, miRNA, and other. Concordant fragments of retrotransposons copies were grouped based on the UCSC RepeatMasker. When computing HMR enrichments over retrotransposon subfamilies only those containing at least 100 members with at least 3 CpGs covered 5-fold were retained. *Long RNA-seq*: 5' coordinates of reads where used for intersection with UCSC tracks and miRBase (Griffiths-Jones 2004) with the following priorities: LINE, LTR, SINE, exon, introns, miRNA, structural RNA and others. *piRNAs*: 5' coordinates of reads where used for intersection with UCSC tracks and miRBase with the following priorities: LINE, LTR, SINE, exon, introns, miRNA, structural RNA and others. Reads corresponding to piRNAs excluded miRNA and structural RNAs.

Average promoter methylation and HMR distribution

LTR copies were retained if they were truncated by at most 10 bases 5' and were at least 300 bases long. LINE copies were retained if they were truncated by at most 50 bases 5' and were at least 500 bases long. Average promoter methylation for comparison between WT and Mili -/- spermatocytes was performed over the same set of LINE and LTR described above with the following additional filtering: at least 3 CpGs covered and at least 30 total CpGs per reads.

L1Md_A promoter alignment

The phylogenetic tree relating L1Md_A insertions in the mouse genome was inferred using clustalw2 on the first 500 bp of each genomic insertion. Insertions were only considered if they were (i) at least 5 kb in length, (ii) had at most 50 bp truncation 5', (iii) contained at least 3 CpG sites in the first 500 bp, and (iv) were covered by at least 20 reads.

qRT-PCR on E13.5 and E16.5 PGCs

An independent replicate of PGCs were sorted directly in Trizol (Invitrogen, also see material and methods for sorting procedure). RNA was extracted and reverse transcribed using Superscript III (Invitrogen) with random hexamer priming. cDNAs were subjected to 40 cycles of qPCR amplification using SYBR master mix (Applied Biosystems) on a realplex4 machine (Eppendorf). Fold changes were computed using the $\Delta\Delta$Ct method and normalizing against the SDHA or MapK1 gene (van den Bergen et al. 2009). L1 primers were from (Ciaudo et al. 2013) and IAP from (Hayashi et al. 2008). SDHA: (F) cacaatctatgaagtgactccttgtt (R) tacctgcgtttcccctcata. Mapk1: (F) ccttcagagcactccagaaagt (R) acaacaccaaaaaggcatcc. L1_ORF2: (F) ggagggacatttcattctcatca (R) gctgctcttgtatttggagcataga.L1Md_T: (F) cagcggtcgccatcttg (R) caccctctcacctgttcagactaa. L1Md_A: (F) ggattccacacgtgatcctaa (R) tcctctatgagcagacctgga. IAPEz_int: (R) atgacgttcagccgcagtatg (F) tttgaggagactgtaccccga

Correlation between expression and promoter methylation

*Genes*: Genes were used if they had RPKM > 0 in at least one of the samples. In each correlation, the top 0.1% expressed genes were removed. Methylation was measured in the +/- 4 kb region around the TSS. Genes were taken from RefSeq with redundancies removed (transcripts collapsed) resulting in ~21k genes. Correlation coefficient was computed using Spearman correlation between methylation and ranked RPKMs. *Long RNA-seq and piRNA from LINEs and LTRs*: Repeat copies were only used if both WT and mutant had at least 3 covered CpG sites and at least 20 total reads in the promoter region. Promoters were defined as 1000 bases internal for LINE and 300 for LTR. Reads from long RNA-seq data were counted anywhere in the full annotated length of the repeat. piRNA reads were counted in the promoter region only.

piRNA enrichment over HMRs, DMRs and default regions

Only the 5' mapping location of reads > 24nt were used. DMRs were defined to include non-overlapping parts of HMRs and must have at least 5 significantly differentially methylated CpGs. Default were all non-HMR or –DMR regions covered.

**Supplementary Tables legends.**

**Table S1.**

Sequencing statistics. This table summarizes the quality of the methylomes presented here. For each sample (spermatocytes WT, Mili -/- replicates and primordial germ cells) conversion rate, average individual CpG coverage, % of CpGs covered at least once and average methylation level genome-wide are shown. The correlation coefficients (Pearson, linear regression) between mutant replicate libraries (Rep1 and Rep2) are shown for the methylation level of all promoters, LINEs and LTRs. As an independent validation, correlations were also computed between our WT library and sperm methylomes generated by (Kobayashi et al. 2012). The total number of sequenced and mapped reads is reported for RNA-seq as well (bottom).

**Table S2.**

Overlap in HMR between WT and Mili -/- spermatocytes. The fraction of WT HMRs overlapping Mili -/- HMRs is reported genome wide and within promoters, CGI (CpG islands) and repeats. The total number and genomic annotations of HMRs and DMRs in each genotype is also reported for promoters, LINE, SINE, LTR or other. WT<Mut denotes hypomethylated DMRs in WT only, whereas Mut<WT denotes the opposite.

**Table S3.**

Summary of the Top 5 sub-families overlapping cHMRs, DMRs and Default regions. For each sub-family, the table quantifies the number of individual genomic insertions that are constitutively hypomethylated (cHMR), hypomethylated in Mili -/- (DMRs) or methylated by default (default).

**Table S4.**

Retrotransposon enrichment within HMRs. For each sub-families, the number of elements overlapping an HMR is reported for each genotype (WT or Mili -/-). The "enrichment" is the ratio of observed/expected. The relative enrichment corresponds to the ratio of enrichment in Mili -/- and WT. The enrichment for sub-families of LTRs (4A) and LINEs (4B) are reported as separate tables.

**Table S5** .

Summary of CpG methylation from (Seisenberger et al. 2012). Low coverage methylation data from Seisenberger et al., (2012) was used to compute the distributions of CpG methylation levels at various stages of germ cell development. The colored heat maps reflect the fraction of elements in each annotation class contained within a given bin of average methylation (in 0.1 increments). Several subfamilies of LINE and LTRs are shown (left and middle section). The distribution is also shown for CpGs within DMRs, cHMRs or genome-wide (right section). Similar tables from this study are also shown for comparison.

**Table S6.**

Summary of long-RNA profiling. The genomic annotation of RNA-seq reads from E13.5 PGCs, E16.5 PGCs, non-PGCs gonadic tissues at E13.5 (Somatic) and WT or Mili-/- spermatocytes are reported. Both absolute read counts (left table) and fractions of total mapped reads (right table) are shown.

**Table S7.**

Expression Tables for LINE SINE and LTRs. The absolute read count, count of reads per million mapped read (RPM) and reads per million mapped reads per kilo-base are shown for all subfamilies of LINE (7A), LTRs (7B) and SINEs (7C). Expression is quantified for E13.5 PGCs, E16.5 PGC, Mili +/- and Mili -/- spermatocytes and non-PGCs gonadic tissues at E13.5 (Somatic) for long-RNAs. Total piRNAs cloned at E13.5, E16.5 as well as MILI and MIWI2 Immuno-precipitated RNAs at E16.5 (MILI-IP or MIWI2-IP) are also shown.

**Table S8.**

Correlation between differential promoter methylation and expression for LINE and LTR subfamilies. For selected subfamilies of LINE and LTR, spearman correlation coefficients were computed between differential promoter methylation and expression from RNA-seq, for PGCs (E13.5 and E16.5), spermatocytes (WT and Mili (-/-)). Promoters were taken as 1000 bases downstream of transcription start sites for LINEs, and 300 bases for LTRs. Differential methylation is defined as average methylation for wild-type minus average methylation for mutant, and copies were not considered unless they had at least 3 CpG sites, and at least 20 mapped reads in each methylome.

**Table S9.**

Genomic annotations of piRNAs. The genomic annotation of piRNA reads from total small RNA libraries cloned at E13.5, E16.5, MILI-IP or MIWI2-IP is shown. Both absolute read counts (left table) and fractions of total mapped reads (right table) are displayed.

**Table S10.**

piRNAs 5'U biases. Nucleotide frequencies for the first 15 bases of all mapped piRNA reads in all libraries.

**Table S11.**

piRNA enrichment over DMRs. The ratio of RPM values for all sequenced piRNA libraries between DMRs and cHMRs or DMRs and default is shown.

**Table S12.**

Correlation between piRNA levels and methylation for retrotransposons. Spearman correlation coefficient between ranked piRNA expression levels and promoter methylation for all copies covered within subfamilies of LINE and LTRs retrotransposon. Correlations are shown for RPMs over the first 1kb of all elements.

## Supplementary references

Ciaudo C, Jay F, Okamoto I, Chen CJ, Sarazin A, Servant N, Barillot E, Heard E, Voinnet O. 2013. RNAi-dependent and independent control of LINE1 accumulation and mobility in mouse embryonic stem cells. *PLoS Genet* **9**: e1003791.

Griffiths-Jones S. 2004. The microRNA Registry. *Nucleic Acids Res* **32**: D109-111.

Hayashi K, Chuva de Sousa Lopes SM, Kaneda M, Tang F, Hajkova P, Lao K, O'Carroll D, Das PP, Tarakhovsky A, Miska EA et al. 2008. MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS One* **3**: e1738.

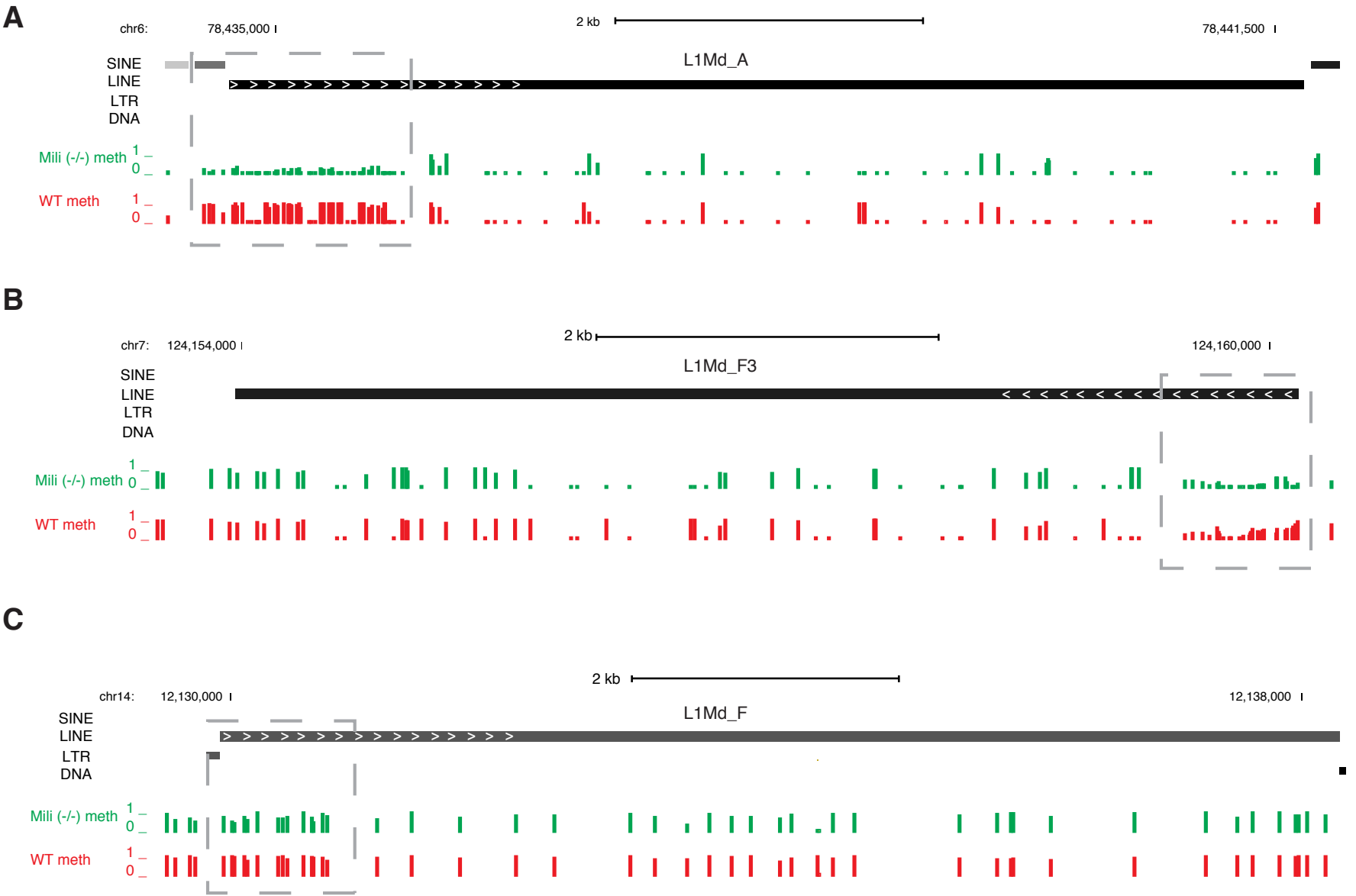
Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, Park J, Butler J, Rafii S, McCombie WR et al. 2011. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell* **44**: 17-28.

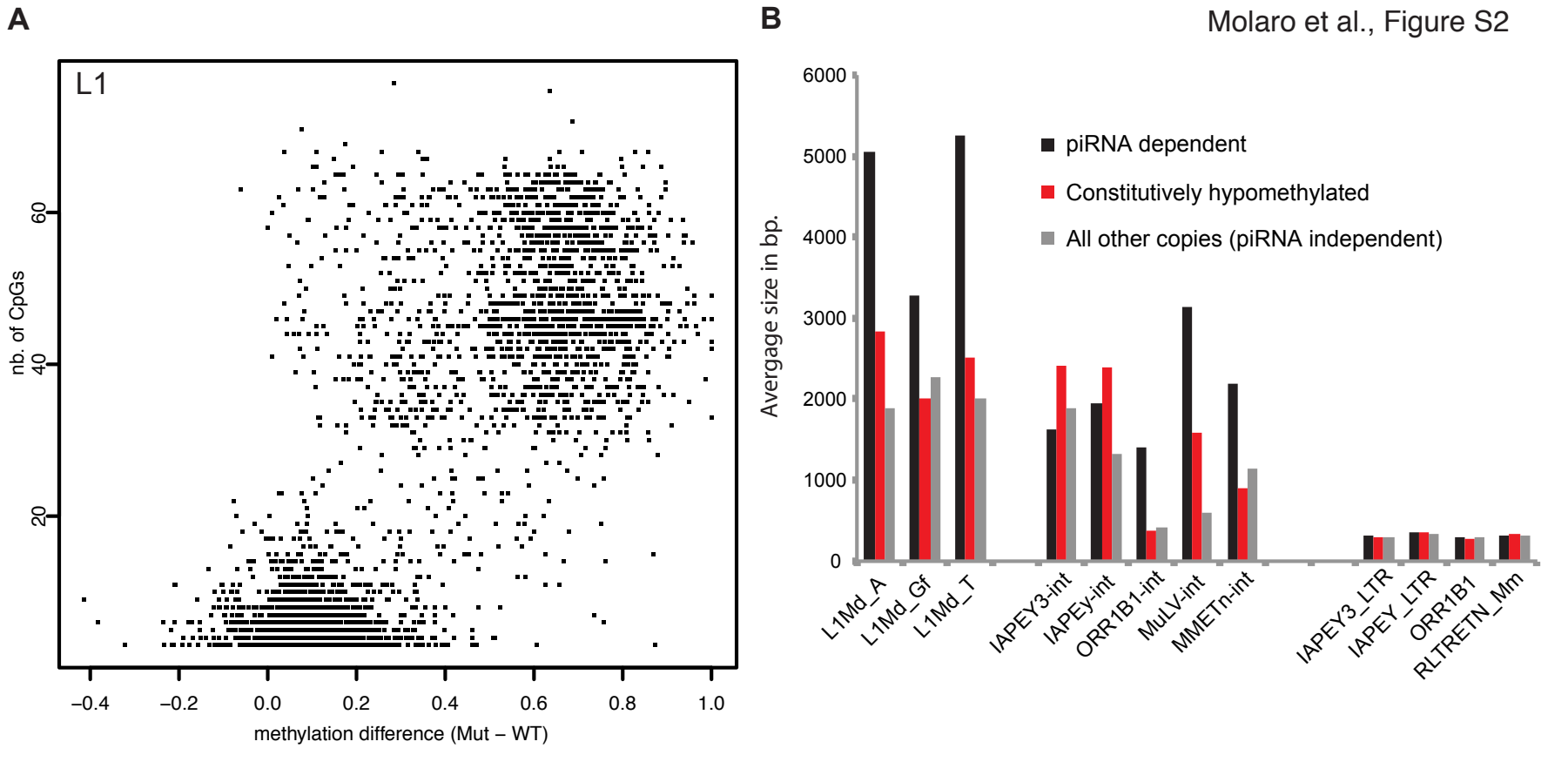
Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y et al. 2012. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* **8**: e1002440.

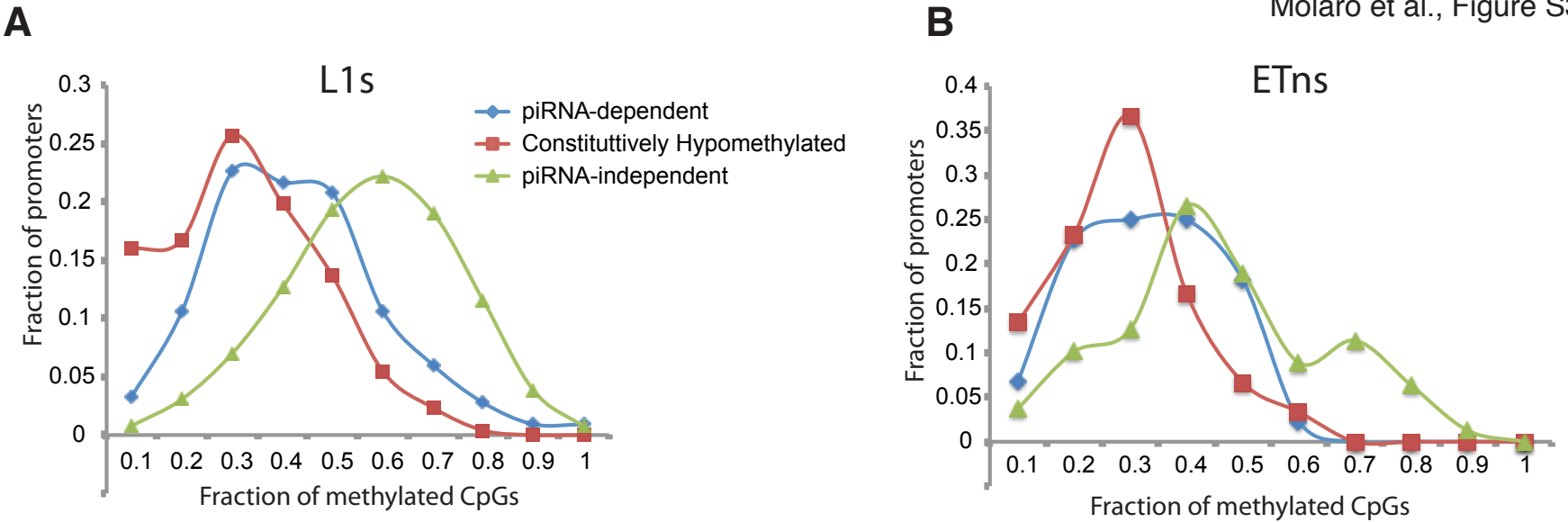
Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. 2011. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**: 1029-1041.

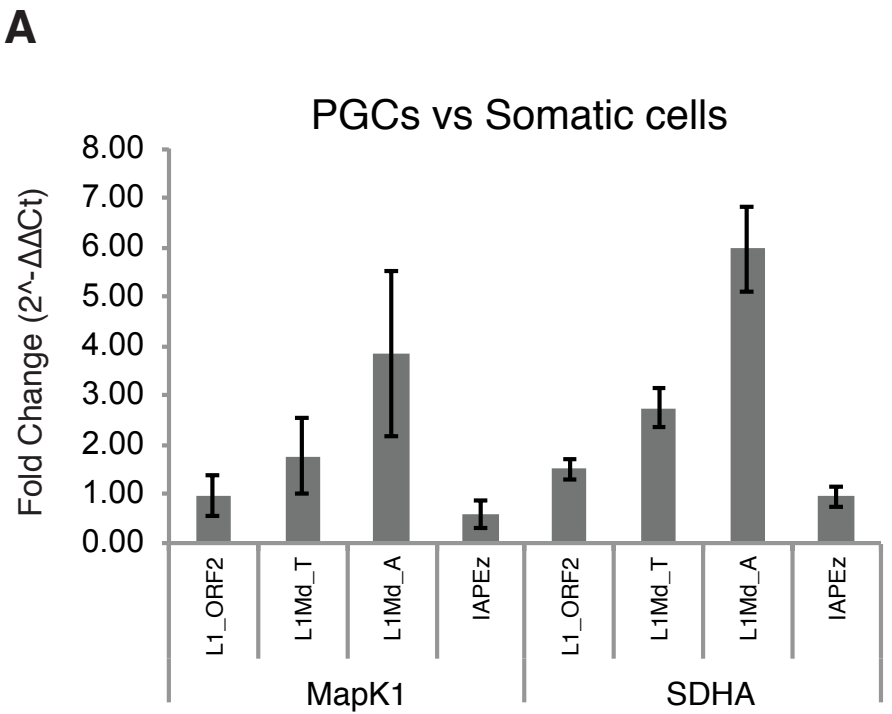
Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W. 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell* **48**: 849-862.

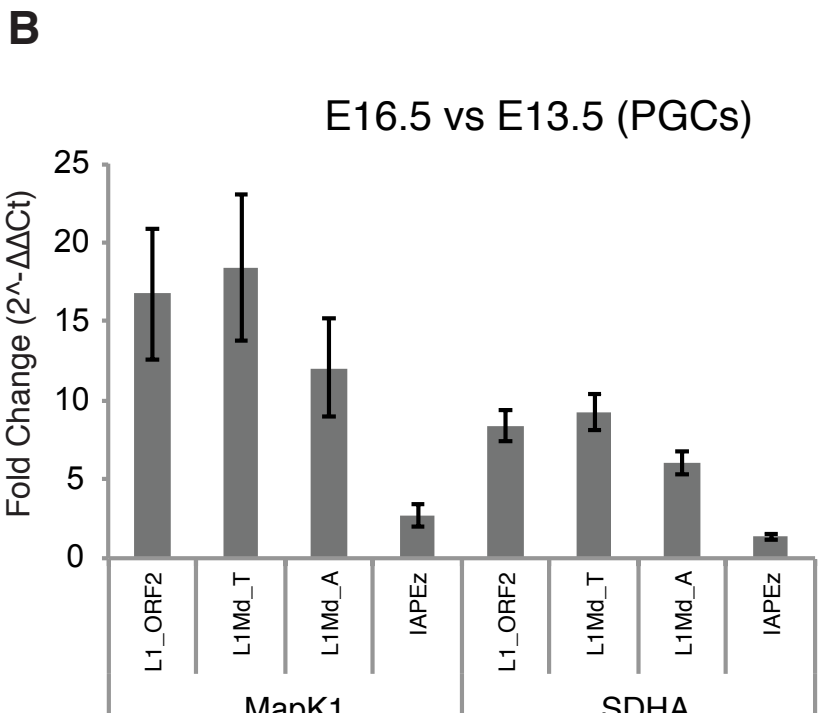
van den Bergen JA, Miles DC, Sinclair AH, Western PS. 2009. Normalizing gene expression levels in mouse fetal germ cells. *Biology of reproduction* **81**: 362-370.

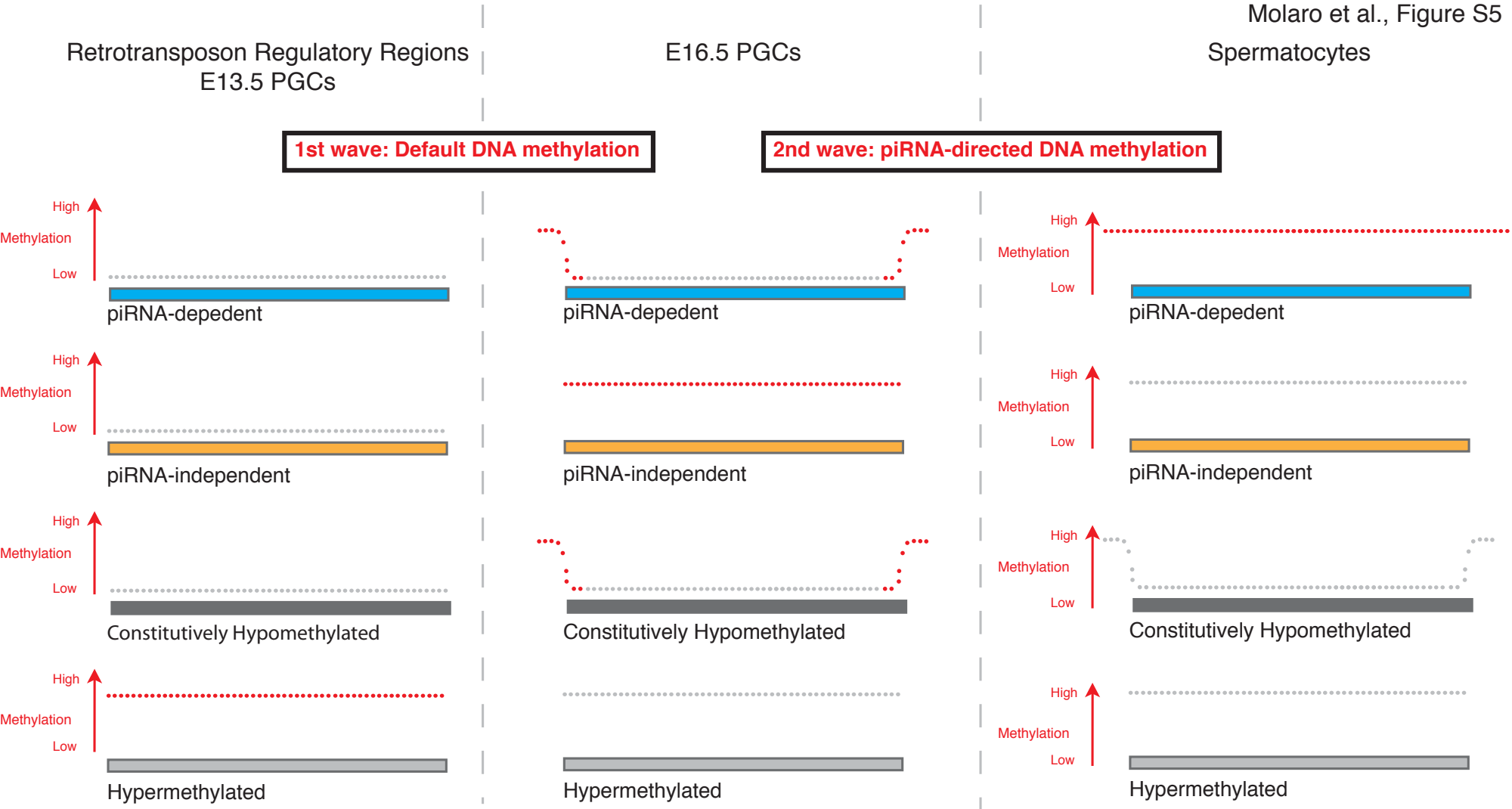

A
chr6:
78,435,000
2 kb
78,441,500
SINE
LINE
LTR
DNA
L1Md_A
Mili (-/-) meth
1
0
WT meth
1
0
B
chr7:
124,154,000
2 kb
124,160,000
SINE
LINE
LTR
DNA
L1Md_F3
Mili (-/-) meth
1
0
WT meth
1
0
C
chr14:
12,130,000
2 kb
12,138,000
SINE
LINE
LTR
DNA
L1Md_F
Mili (-/-) meth
1
0
WT meth
1
0


A
L1
nb. of CpGs
20
40
60
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
methylation difference (Mut – WT)
B
Molaro et al., Figure S2
Avergage size in bp.
0
1000
2000
3000
4000
5000
6000
piRNA dependent
Constitutively hypomethylated
All other copies (piRNA independent)
L1Md_A
L1Md_Gf
L1Md_T
IAPEY3-int
IAPEy-int
ORR1B1-int
MuLV-int
MMETn-int
IAPEY3_LTR
IAPEY_LTR
ORR1B1
RLTRETN_Mm


A
L1s
piRNA-dependent
Constituttively Hypomethylated
piRNA-independent
Fraction of promoters
0.3
0.25
0.2
0.15
0.1
0.05
0
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Fraction of methylated CpGs
B
ETns
Fraction of promoters
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Fraction of methylated CpGs


A
PGCs vs Somatic cells
Fold Change (2^-ΔΔCt)
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
L1_ORF2
L1Md_T
L1Md_A
IAPEz
L1_ORF2
L1Md_T
L1Md_A
IAPEz
MapK1
SDHA


B
E16.5 vs E13.5 (PGCs)
Fold Change (2^-ΔΔCt)
25
20
15
10
5
0
L1_ORF2
L1Md_T
L1Md_A
IAPEz
L1_ORF2
L1Md_T
L1Md_A
IAPEz
MapK1
SDHA


Retrotransposon Regulatory Regions
E13.5 PGCs
E16.5 PGCs
Spermatocytes
1st wave: Default DNA methylation
2nd wave: piRNA-directed DNA methylation
High
Methylation
Low
piRNA-depedent
piRNA-independent
Constitutively Hypomethylated
Hypermethylated